



# Malaria Genetics and Genomics

National Center for Biotechnology Information ■ National Library of Medicine ■ National Institutes of Health ■ Department of Health and Human Services

*Anopheles gambiae* (African malaria mosquito) is the primary vector that transmits *Plasmodium falciparum*, the causative agent of human malaria. NCBI provides direct access to the genomic sequence data of *A. gambiae*, *P. falciparum* and other malaria-associated genomic resources described below and featured at [www.ncbi.nlm.nih.gov/projects/Malaria/](http://www.ncbi.nlm.nih.gov/projects/Malaria/)



## Content of Entrez Databases

**Nucleotide:** Genomic DNA and cDNA sequences from GenBank and NCBI Reference Sequences.

**Protein:** Protein sequences from PIR, PRF, SWISS-PROT, PDB, conceptual translations of GenBank sequences and NCBI Reference Sequences.

**Structure:** Experimentally determined 3D structures from Protein Data Bank (PDB).

**Genomes:** Finished genomic sequences for *P. falciparum* (NC\_#####) and mitochondrial records for *A. gambiae* and *H. sapiens*.

**SNP:** DNA polymorphisms (SNPs, insertions, deletions, microsatellites, etc.).

**UniGene:** Gene-oriented clusters of expressed sequence tags (ESTs) and mRNA sequences.

**UniSTS:** A database of Sequence tagged sites.

**PopSet:** Global alignments of DNA sequences from population studies.

## Reference Sequences [www.ncbi.nlm.nih.gov/LocusLink/refseq.html](http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html)

NCBI Reference Sequences (RefSeqs) are curated, non-redundant sequences derived from the primary sequence data in GenBank. RefSeqs include curated mRNA (NM\_#####) and protein (NP\_#####) sequences, as well as those of whole chromosomes (NC\_#####), among others. For *P. falciparum*, NC records in Entrez Genomes represent the finished chromosomes and were assembled from GenBank sequences submitted by the Malaria Genome Project (MGP) Consortium. The many NP records were derived from the coding sequence annotations provided on the relevant GenBank sequences. For *A. gambiae* the mitochondrial genome is available as an NC record, along with a small number of NP records.

## dbEST [www.ncbi.nlm.nih.gov/dbEST/index.html](http://www.ncbi.nlm.nih.gov/dbEST/index.html)

A bulk sequence division of GenBank containing ESTs that represent short (200-500 bp), single-read sequence data from expressed genes in a cell, tissue, or organ under specific cellular conditions.

## Trace Archive [www.ncbi.nlm.nih.gov/Traces/trace.cgi?](http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?)

A repository of raw sequence data from sequencing centers.

## Indexing of *A. gambiae* and *P. falciparum* Genomic Sequences

*A. gambiae*: The genome sequencing project has accession **AAAB00000000**, with the first version of the assembly having accession **AAAB01000000** containing scaffolds **AAAB01000001-AAAB01069724**.

*P. falciparum*: The individual chromosomes are each represented by one NC Reference Sequence.

## Data Access Using Entrez: Entrez Queries

From any Entrez database:

*Anopheles gambiae*[organism]  
*Plasmodium falciparum*[organism]  
 Human[organism]

For Reference Sequences (RefSeqs) in Entrez Nucleotide or Protein:

*Plasmodium falciparum*[organism] AND srcdb refseq  
 [properties]

For ESTs in Entrez Nucleotide:

*Anopheles gambiae*[organism] AND gbdiv est[properties]

## Data Access Using FTP

<ftp://ftp.ncbi.nih.gov/pub/Malaria/>  
[ftp://ftp.ncbi.nih.gov/genomes/Anopheles\\_gambiae/](ftp://ftp.ncbi.nih.gov/genomes/Anopheles_gambiae/)  
[ftp://ftp.ncbi.nih.gov/genomes/P\\_falciparum/](ftp://ftp.ncbi.nih.gov/genomes/P_falciparum/)  
[ftp://ftp.ncbi.nih.gov/genomes/H\\_sapiens](ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/)

## Data Records in Selected Databases Held at NCBI as of November 2002

	<i>A. gambiae</i>	<i>P. falciparum</i>	<i>H. sapiens</i>
<b>Entrez</b>			
Nucleotide	225,109	25,278	6,229,582
Protein	15,869	14,822	176,075
Structure	1	18	3,768
Genomes	1	15	1
SNP	398,887	0	2,703,997
UniGene	2,568	0	104,214
UniSTS	0	863	186,587
PopSet	15	19	199
<b>Ref Seq</b>			
Nucleotide	1	15	41,658
Protein	13	5,270	39,580
<b>dbEST</b>	97,032	18,742	4,811,971
<b>Trace Archive</b>	4,888,159	0	9,897,257
<b>Genomic Assembly</b>	yes	yes	yes

***P. falciparum***

The NCBI Map Viewer for *P. falciparum* displays genetic information for all 14 chromosomes using the following four maps:

**Chromosome:** displays the NC Reference Sequence record for the chromosome.

**Component:** displays the tiling path of GenBank sequences used to assemble the NC record.

**Gene:** displays known or putative genes based on alignments of the NP records to the NC records using blastx.

**Genetic:** displays the genetic linkage map for *P. falciparum* and relevant STS markers.

**NEW** **Py Proteins:** displays *P. yoelii* proteins aligned to *P. falciparum* chromosomes.

[www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map\\_search?chr=pfalciparum.inf](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map_search?chr=pfalciparum.inf)

***A. gambiae***

The NCBI Map Viewer for *A. gambiae* displays the following cytogenetic and sequence maps, along with an alignment of *D. melanogaster* proteins:

**Satellites:** sequence map of satellite positions.

**Satellites on Bands:** cytogenetic map of satellite positions.

**Bands:** ideogram of the chromosome with cytogenetic coordinates.

**Scaffolds on Bands:** cytogenetic map of scaffold positions.

**Scaffolds:** sequence map of scaffold positions.

**NEW** **Dm proteins:** sequence map of *D. melanogaster* proteins aligned to *A. gambiae* scaffolds.

**Gene:** sequence map of annotated *A. gambiae* genes with links to protein sequences.

[www.ncbi.nlm.nih.gov/mapview/map\\_search.cgi?chr=agambiae.inf](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=agambiae.inf)

***A. gambiae***

NCBI provides three ways to perform BLAST searches against *A. gambiae* genomic data:

**Genomic BLAST:** provides BLAST searches against *A. gambiae* genomic scaffolds or proteins. The available versions of BLAST are blastp, blastn, blastx, tblastn, and Megablast.

[www.ncbi.nlm.nih.gov/PMGifs/Genomes/agambiae.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/agambiae.html)

**Trace Archive BLAST:** allows BLAST searches against the raw trace data for *A. gambiae* using Megablast, and is optimized for searching with *A. gambiae* queries.

[www.ncbi.nlm.nih.gov/blast/mmtrace.html](http://www.ncbi.nlm.nih.gov/blast/mmtrace.html)

**Cross-species Trace Archive BLAST:** allows BLAST searches against the raw trace data for *A. gambiae* using discontinuous Megablast, and is optimized for searching with queries from other species.

[www.ncbi.nlm.nih.gov/blast/tracemb.html](http://www.ncbi.nlm.nih.gov/blast/tracemb.html)

***P. falciparum***

NCBI provides three ways to perform BLAST searches against *P. falciparum* genomic data, as well as genomic data for other *Plasmodium* species:

**Genomic BLAST:** provides access to both DNA and protein databases for *P. falciparum*, *P. yoelii*, or both organisms. The available versions of BLAST are blastp, blastn, blastx, tblastn, and Megablast.

[www.ncbi.nlm.nih.gov/PMGifs/Genomes/plasmodium.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/plasmodium.html)

**GenBank BLAST:** enables searches against selected GenBank divisions (EST, GSS, STS, and HTGS), along with the standard nr database, or a combination of all of the above. Searches can be restricted to *P. falciparum*, all *Plasmodium*, or all toxoplasma. Only nucleotide databases are available here, so the BLAST versions provided are blastn, tblastn, and tblastx.

[www.ncbi.nlm.nih.gov/projects/Malaria/plasmodiumbl.html](http://www.ncbi.nlm.nih.gov/projects/Malaria/plasmodiumbl.html)

**Custom BLAST:** provides BLAST access to unfinished sequences provided by the MGP Consortium but not necessarily in GenBank. Databases for individual chromosomes are available for *P. falciparum*, while broader databases are available for *P. berghei*, *P. vivax*, *P. chabaudi*, *P. knowlesi*, and *P. yoelii*.

[www.ncbi.nlm.nih.gov/projects/Malaria/plasmodiumblcus.html](http://www.ncbi.nlm.nih.gov/projects/Malaria/plasmodiumblcus.html)

**Other Useful NCBI Resources**

**Rodent Malaria Resources:** This page provides data and information relevant to the four species of *Plasmodium* that cause malaria in rodents: *P. berghei*, *P. chabaudi*, *P. yoelii* and *P. vinckei*.

[www.ncbi.nlm.nih.gov/projects/Malaria/Rodent/index.html](http://www.ncbi.nlm.nih.gov/projects/Malaria/Rodent/index.html)

**PubMed:** Access to the Biomedical Literature. PubMed, a service of the National Library of Medicine, provides free, on-line access to over twelve million abstracts indexed in MEDLINE. Key terms in each abstract are hyperlinked to Entrez Books for additional information content. When available, free electronic versions of published reports can be downloaded from PubMed Central. LinkOut provides links to related journal entries and additional resources within and outside of NCBI.

[www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed)

